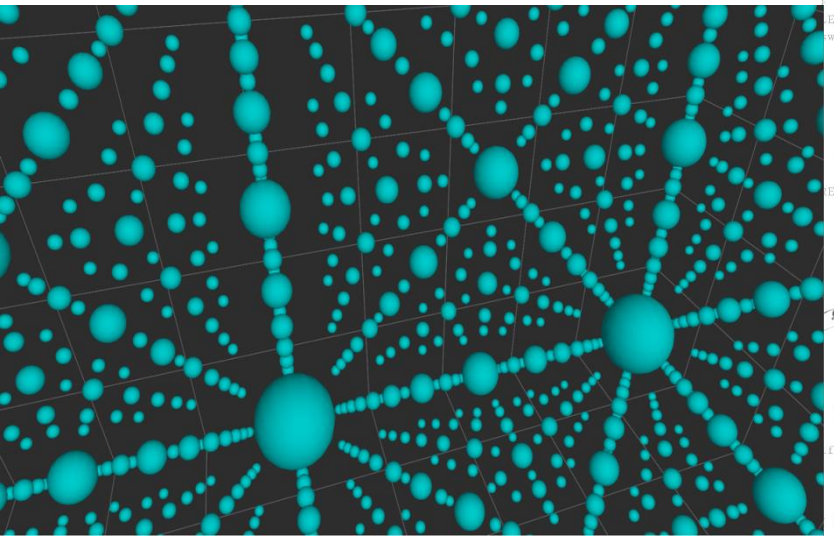
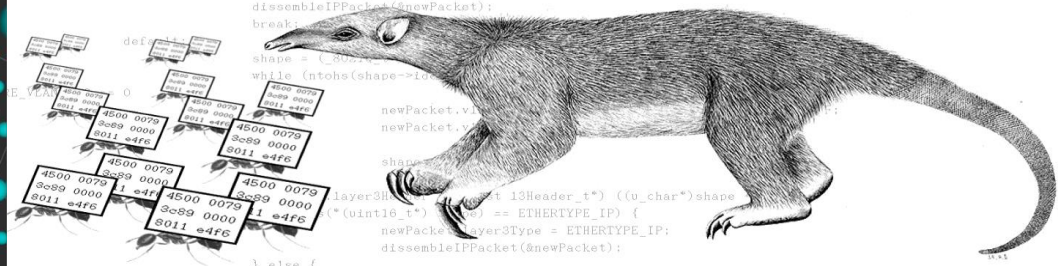
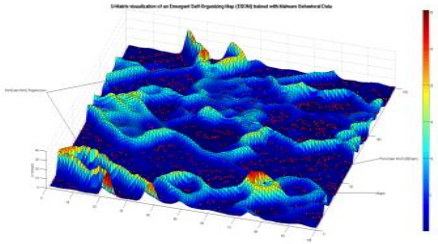
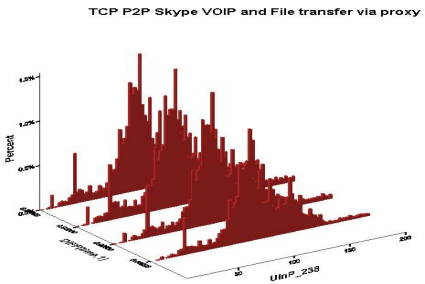
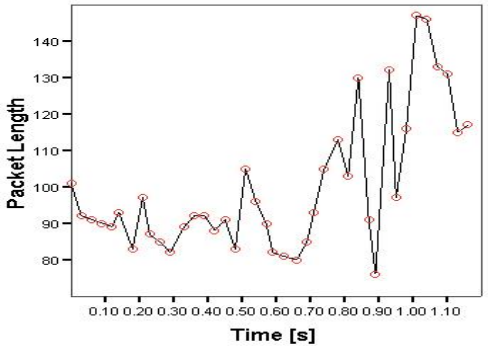
Troubleshooting & Traffic Mining(TM)









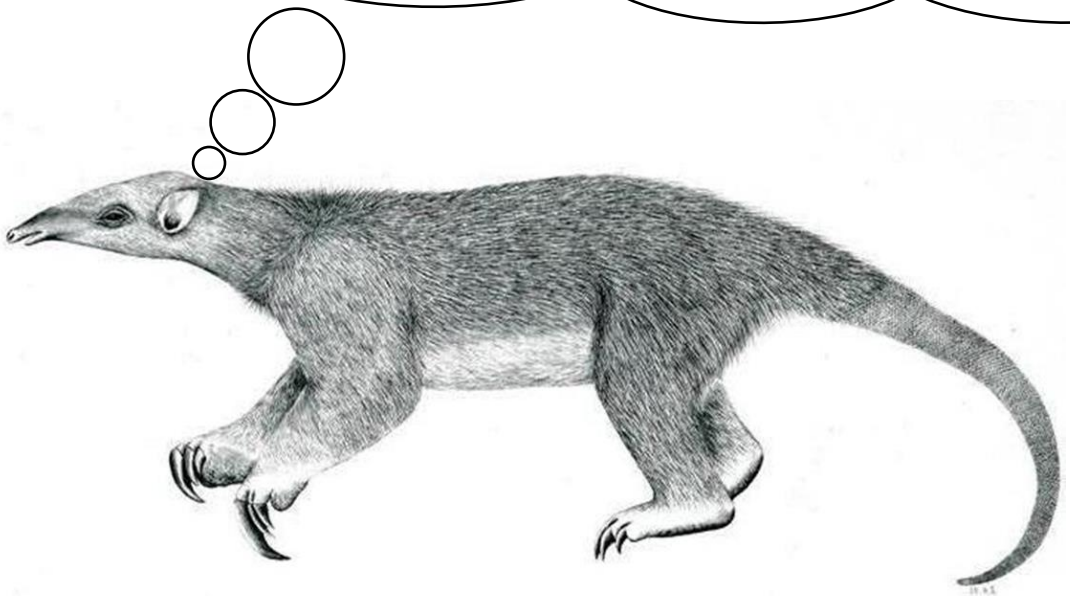


Stefan Burschka

Together ahead. RUAG

Agenda

- Intro: Who, What, Problems with big traffic data
- How to address problems with brain & tools
- Lots of examples and exercises



Together ahead. RUAG

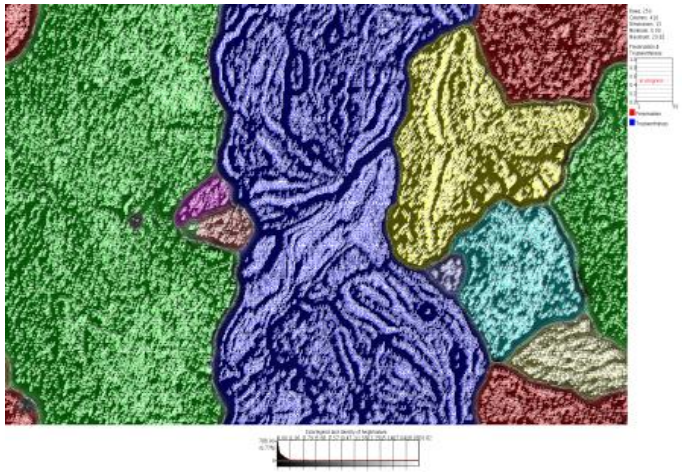
What we do:

Products: Network Troubleshooting, Forensics, Security

- TRANALYZER(T3): High Speed and Volume Traffic Analyzer
- TRAVIZ3: Graphical Toolset for Tranalyzer
- Complete Tool Sets for TM & Forensics
- Artificial Intelligence Plugins

Research:

- Brain support 4 multi-dimensional datasets
- Encrypted TM
- Big Data Visualization (Traviz)
- Malware and covert channel detection
- Nifty stuff



Together ahead. RUAG

Data (Traffic) Mining ← → Forensics

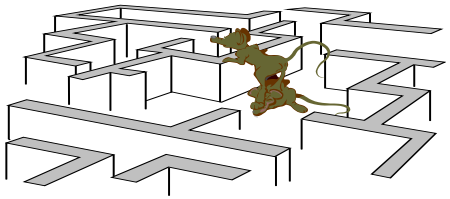
Input: Huge — Tiny Answer
Output: Tiny — Large Puzzle

Almost Everything

→ Find what everybody missed
Find patterns, anomalies and the undetectable.

Almost Nothing

193.5.230.129, 80



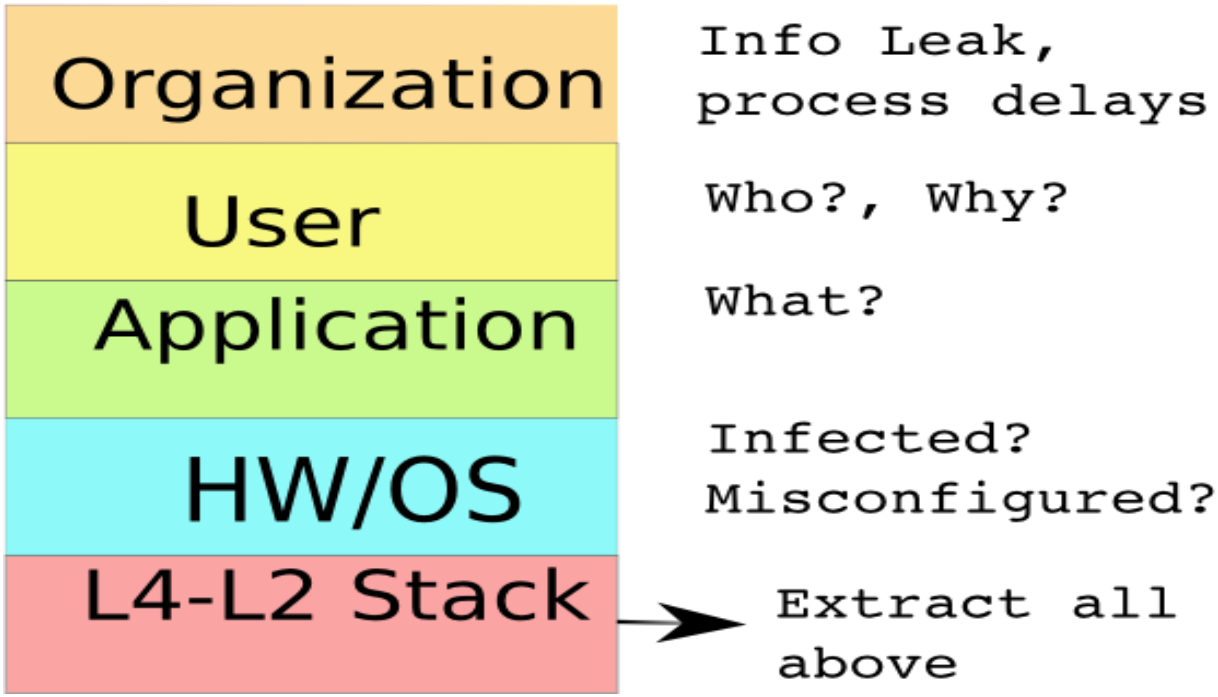
→ Tell me everything and especially
Owner, gender, colour of hair and underpants, talks to pigs?, OS, type of computer ...

Traffic Mining(TM):

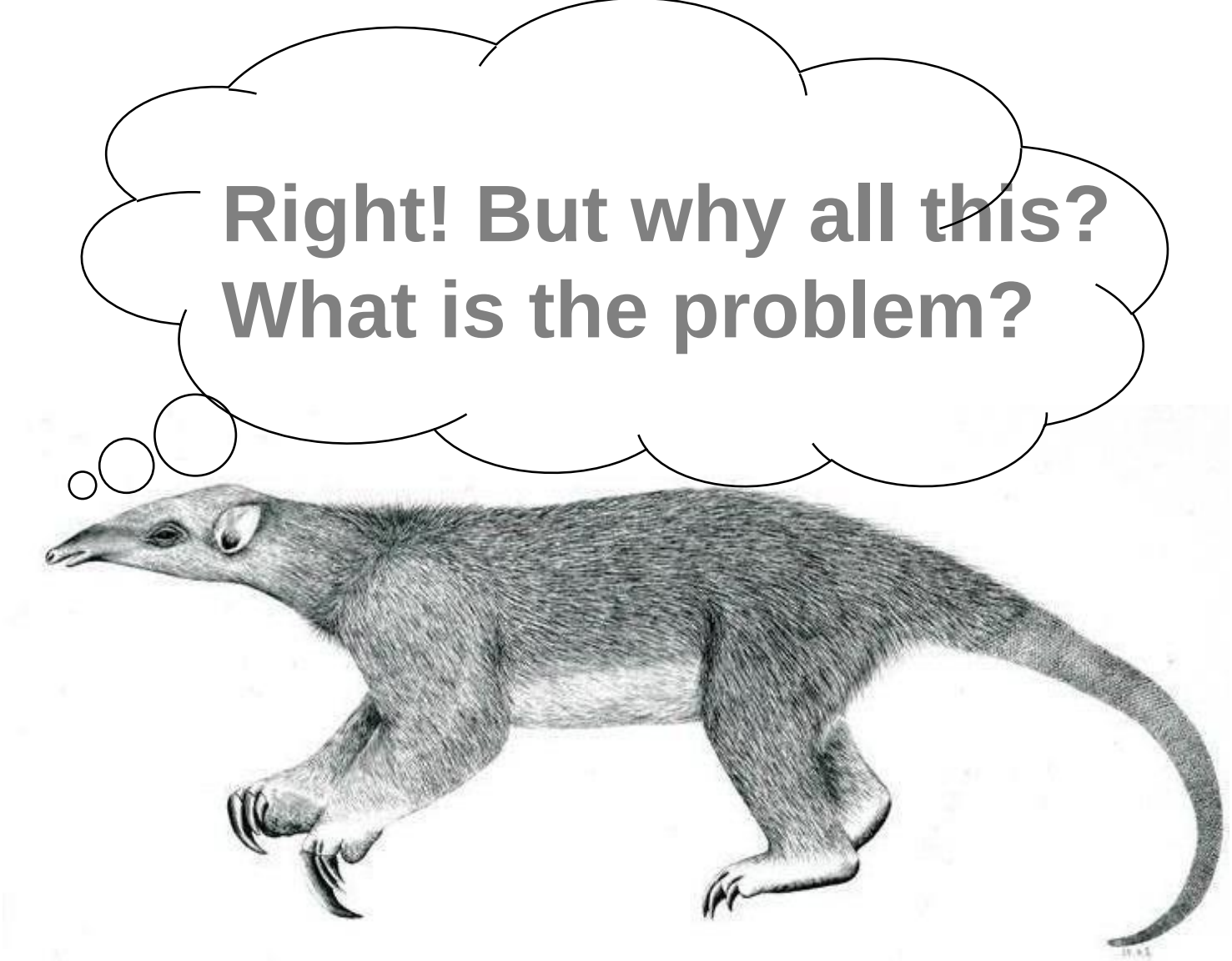
Hidden Knowledge: Listen | See, Understand, Invariants, Model

Application in

- Troubleshooting, Security (Classification, Encrypted TM)
- Netzwerk usage (VoiP, P2P traffic shaping, application/user profiling)
- Profiling & Marketing (usage performance- & market- index)
- Law enforcement and Legal Interception (Indication/Evidence)



Together ahead. RUAG



Together ahead. RUAG

Problem: Controlled Flight into Terrain

- There is a union for the lifeform SW
- Seldom problem oriented KISS Engineering
- High Complexity, small fault tolerance

- Time- and economic pressure, info overload

- Banana Principle: Conditioning of cutomers
- SW became a weapon

The Network is slow, The Network is insecure;
NO, it's not Microsoft, shut up, It wasn't me ...

Manager (MBA)
Always right, DoR
License to Powerpoint

Production (poor Techie)
Knows, Always warned,
Always his fault: FUBAR
License to get fired

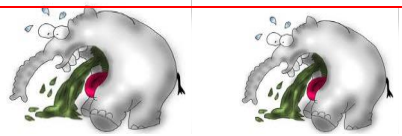
Finance (MBA)
Knows basic calculus
License to Excel

We didn't find the problem in 4 months, can you do the job in 2 weeks? (We supply only 11TB data)

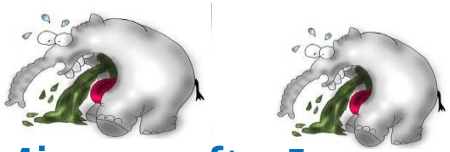
Together ahead. RUAG

Large unstructured Traffic Datasets(/s)

- Problem Comprehension? WTF am I looking for?
- Data Selection: I'd like to have … ←→ Eat what you get!
- Preprocessing: Dimension Reduction, Feature Selection
- Data Integrity: Churchill / Murphy



- Formats: XML, EXCEL ←→ Binary, (txt)



- Tools/Interfaces: Microsoft, Java ←→  Bash, awk,DB
- Operating Storage: Cache ←→ Memory ←→ SSD ←→ Disk

- FBHDD Visualization, AI

Together ahead. RUAG

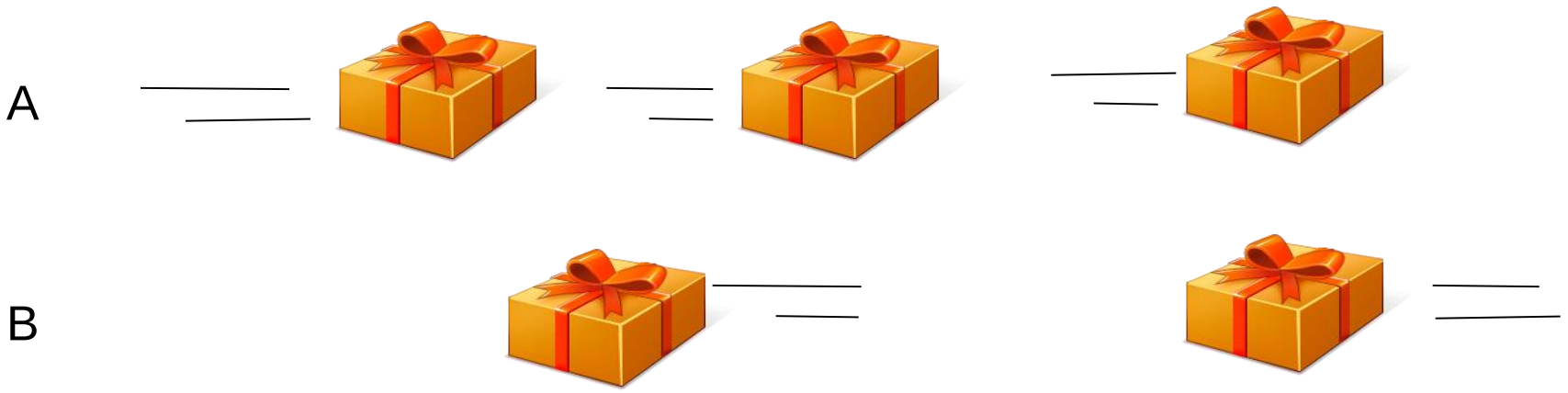
Exercise: What is wrong here?

Together ahead. RUAG

See the disaster now?
Now you have context!

Together ahead. RUAG

Preprocessing/Context/Dimension Reduction Versatile Flow Compression



Definition: (6-Tuple)

Vlan(s), srcIP, srcpPort, dstIP, dstPort, L4Protocol

Or why not a bit more context and meaning ?

srcWho, dstWho

srcNetwork, dstNetwork

Bad, Good

Internal / External

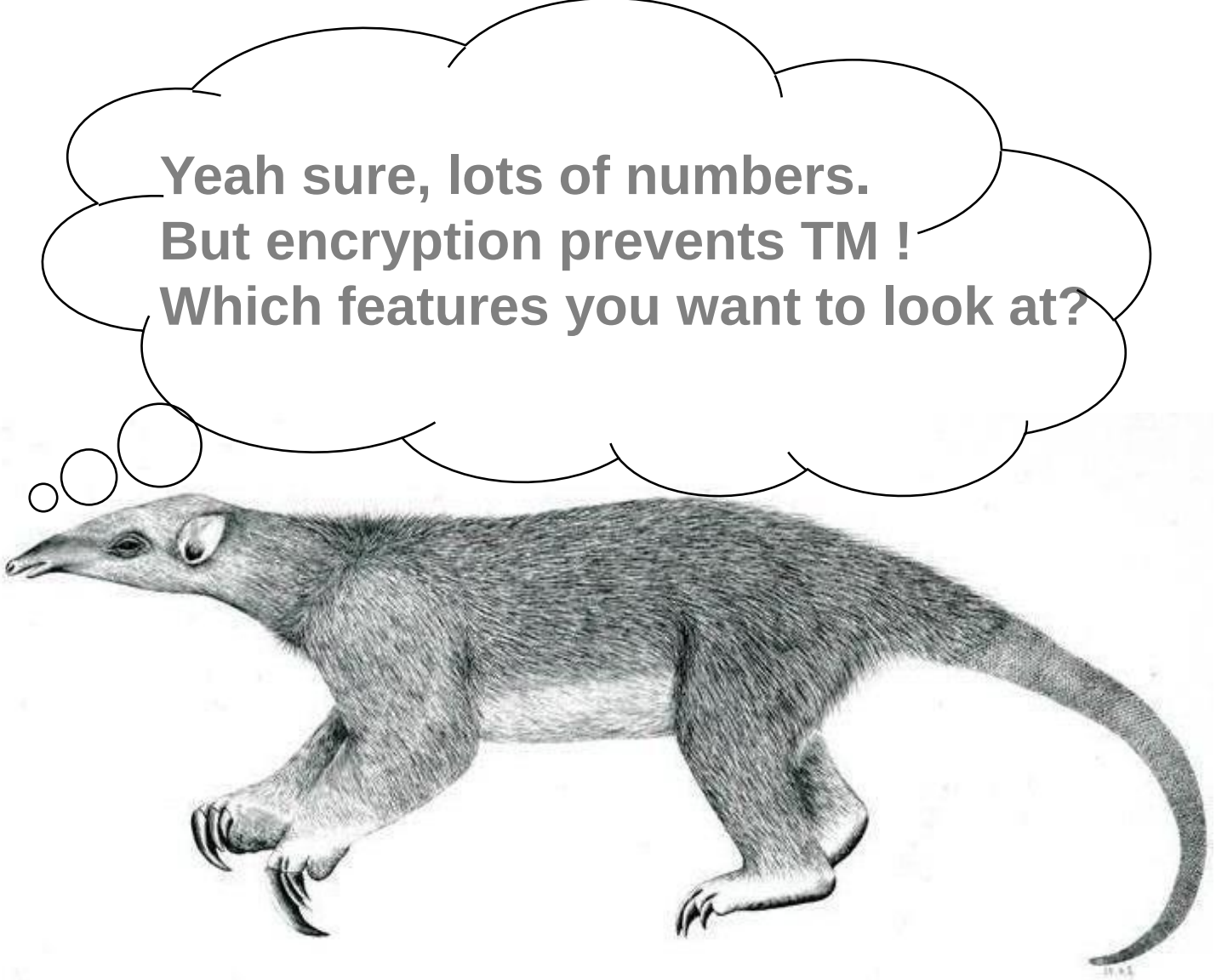
Together ahead. RUAG

Tranalyzer Flow Example

```
A       1196278772.439355 1196279184.642073 412.202718   0x9B42       22
192.168.1.10 0x00000001   2119 182.236.128.34 0x8008036C
80      6       00:0f:1f:cf:7c:45_00:00:0c:07:ac:0a_6387        http    6387    8272
464     5437587         0       579     15.494803       1.125660        -0.128590       -
0.999829        1       87      128     128     0x00    0x42    0x0000          116     464
6231    4116    5437724                 2253    63754   64831.988281            62501   65535
3342    2904    5713    0x18    0xF900          0x0000          0x03    0x00000000
0x0000          -1.0    3       36578   1       ...

B       1196278772.409312 1196279184.642073 412.232761   0x9B43       22
 182.236.128.34       0x8008036C  80      192.168.1.10  0x00000001
2119    6       00:d0:00:64:d0:00_00:0f:1f:cf:7c:45_8272        http    8272    6387
5437587         464     0       1380    20.066333       13190.574633            0.128590
0.999829        1       3       63      63      0x00    0x42    0x0000          8146
        5440245                 109     116     464     8104    5840    5840.000000     65535
0       0       0       0       0x18    0x1B00          0x0000          0x03    0x00000000
0x0000          -1.0    36578   3       1       ...
```

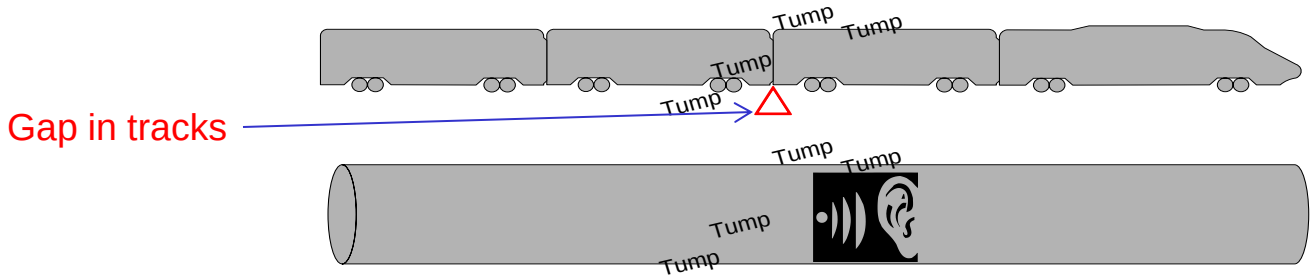
Together ahead. RUAG



Together ahead. RUAG

Encrypted TM: Packet Length Magic

Distinguish  from  by listening



Sound ~ $\vec{F} = d\vec{p}/dt = dm/dt \cdot \vec{v} + m \cdot d\vec{v}/dt$

$$dm/dt = dm/dpkt \cdot dpkt/dt$$

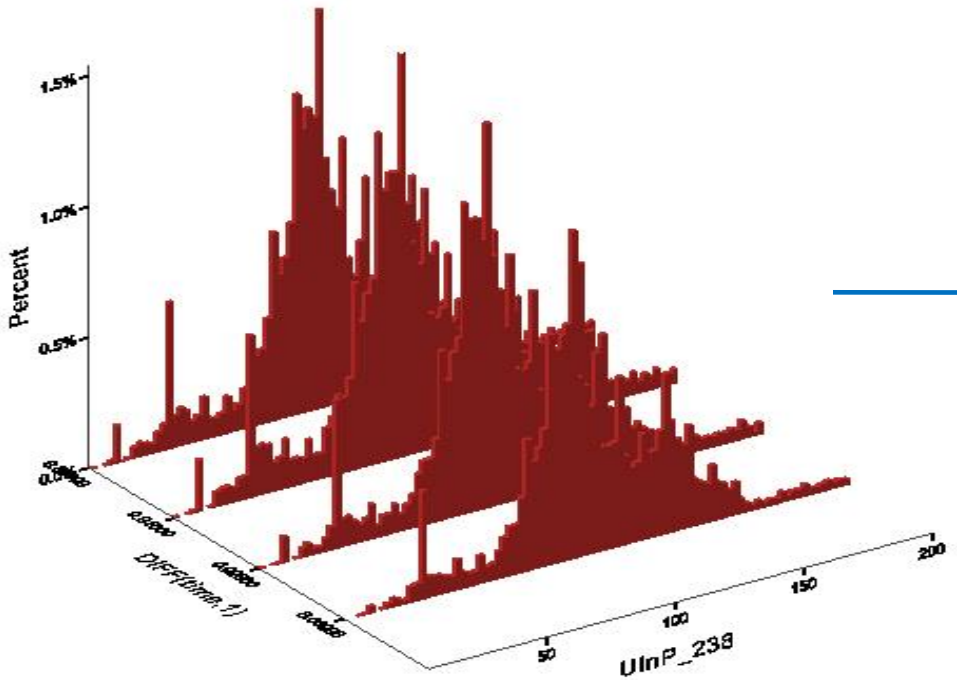
Packet Length

Packet Fire Rate (Interdistance)

Together ahead. RUAG

3D Statistical Application / User profiling

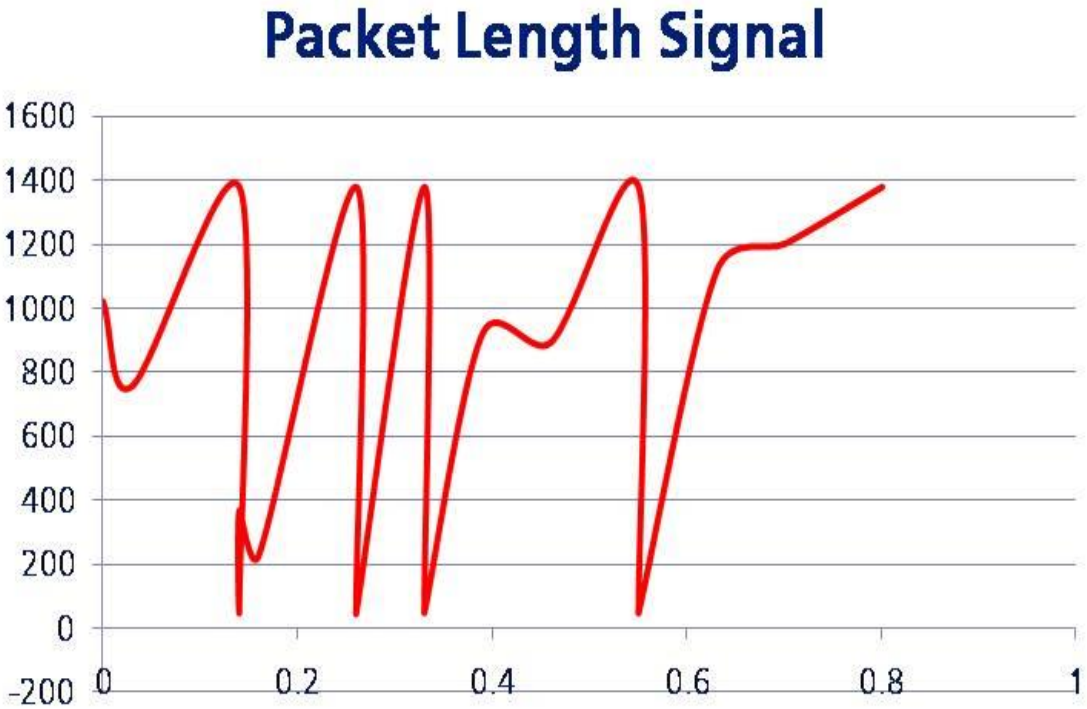
Packet length-Interdistance Statistics: Fingerprinting

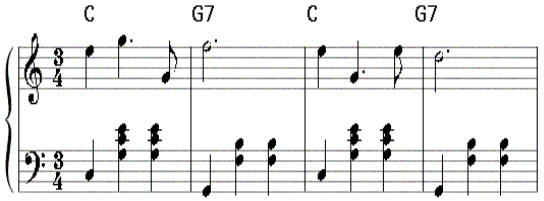


→ Vulnerable against TM

Together ahead. RUAG

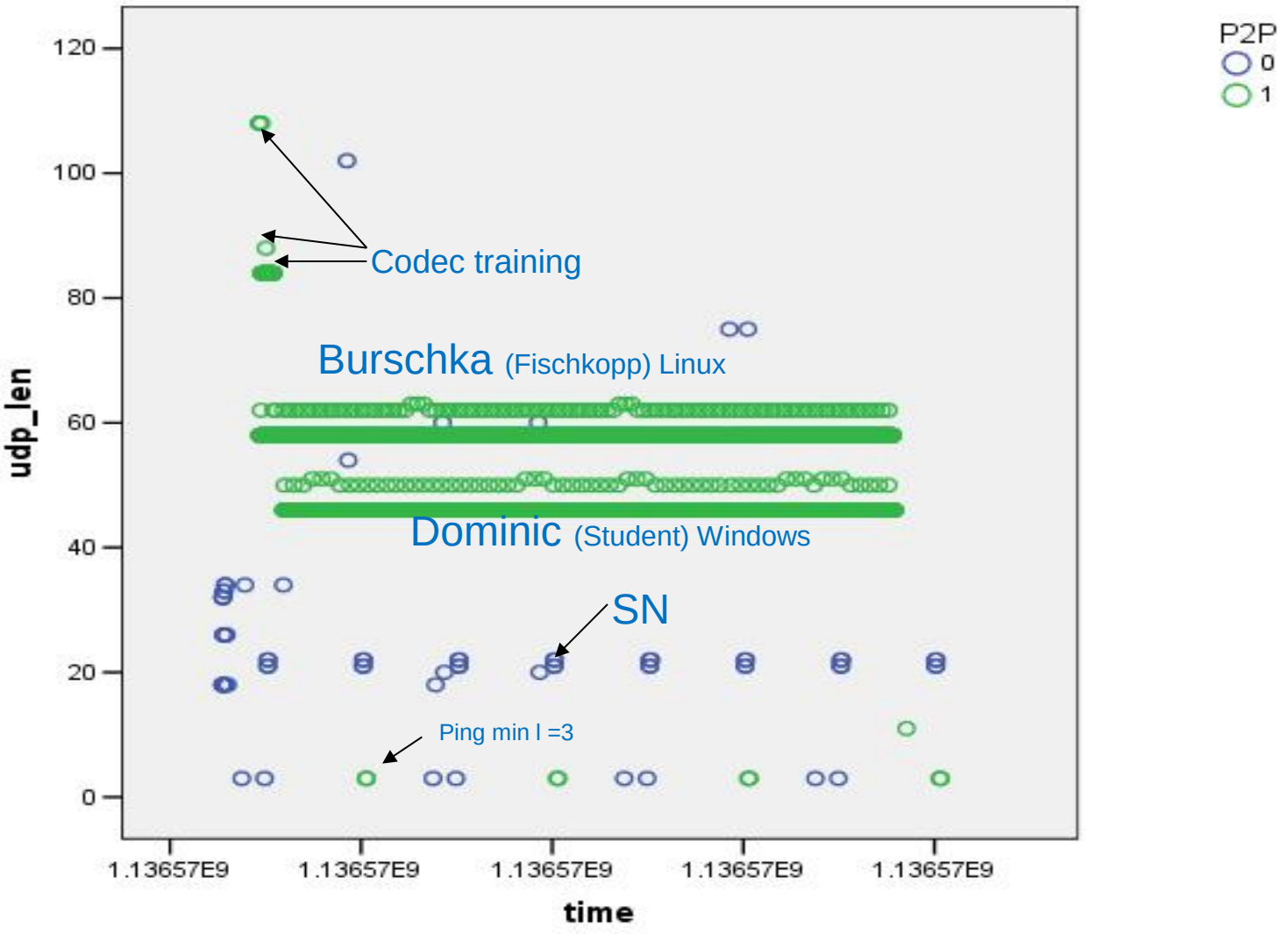
Packet Signal: Encrypted VoIP Mining

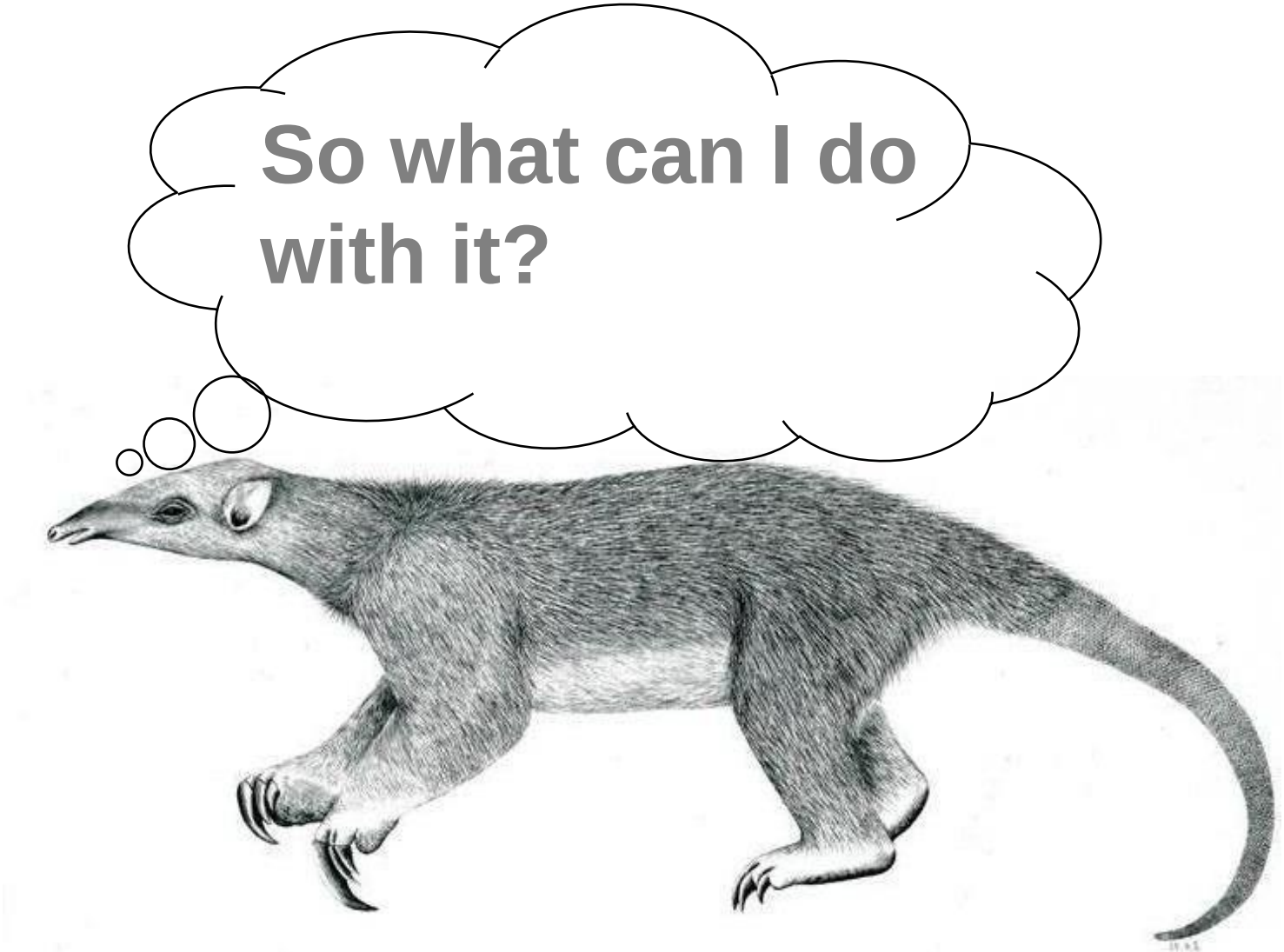




Together ahead. RUAG

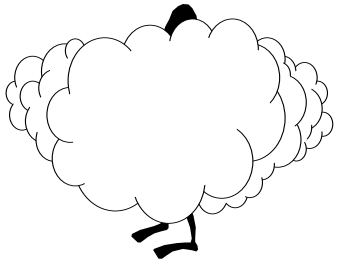
Exercise: Multiple Flow Packet Length Signal See the features?



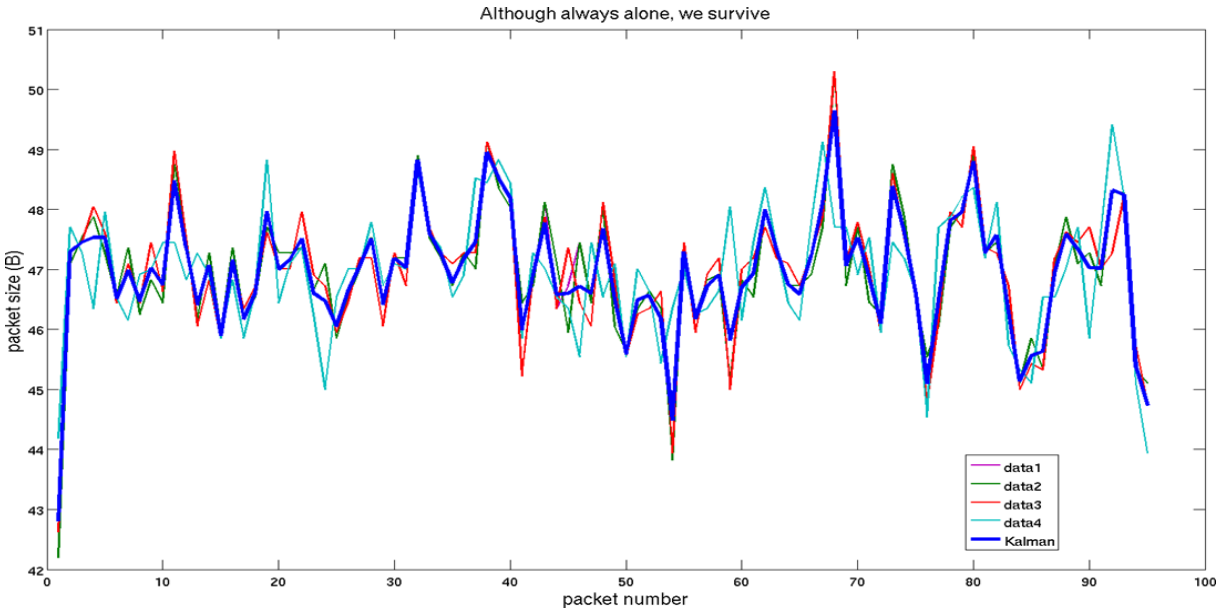


Together ahead. RUAG

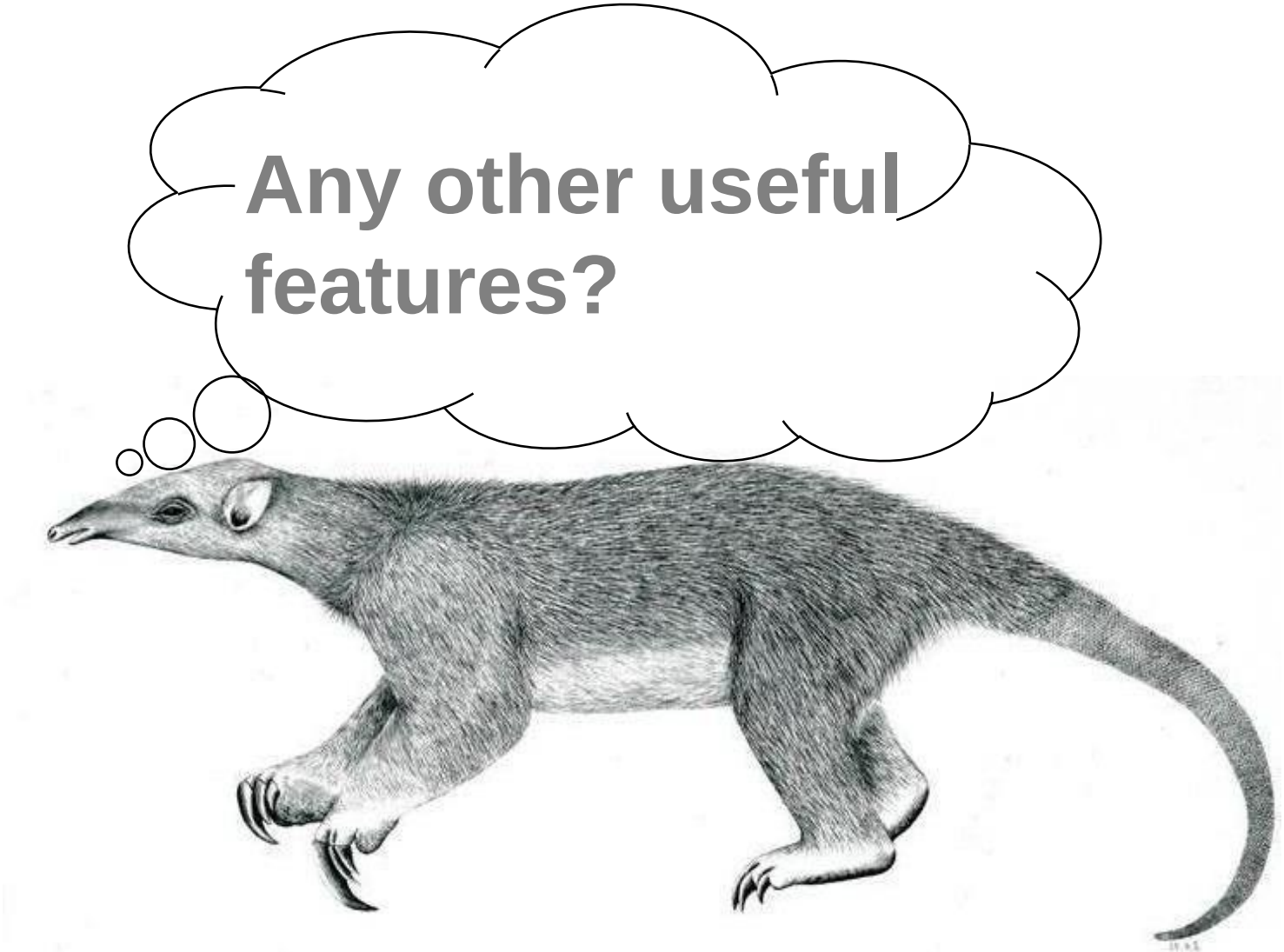
Encryption TM



- SSH Command & Parameter Estimation
- IPSEC Tunnel: Application / User Profiling
- Skype Content Guessing: CCC 2011 (Datamining for Hackers)

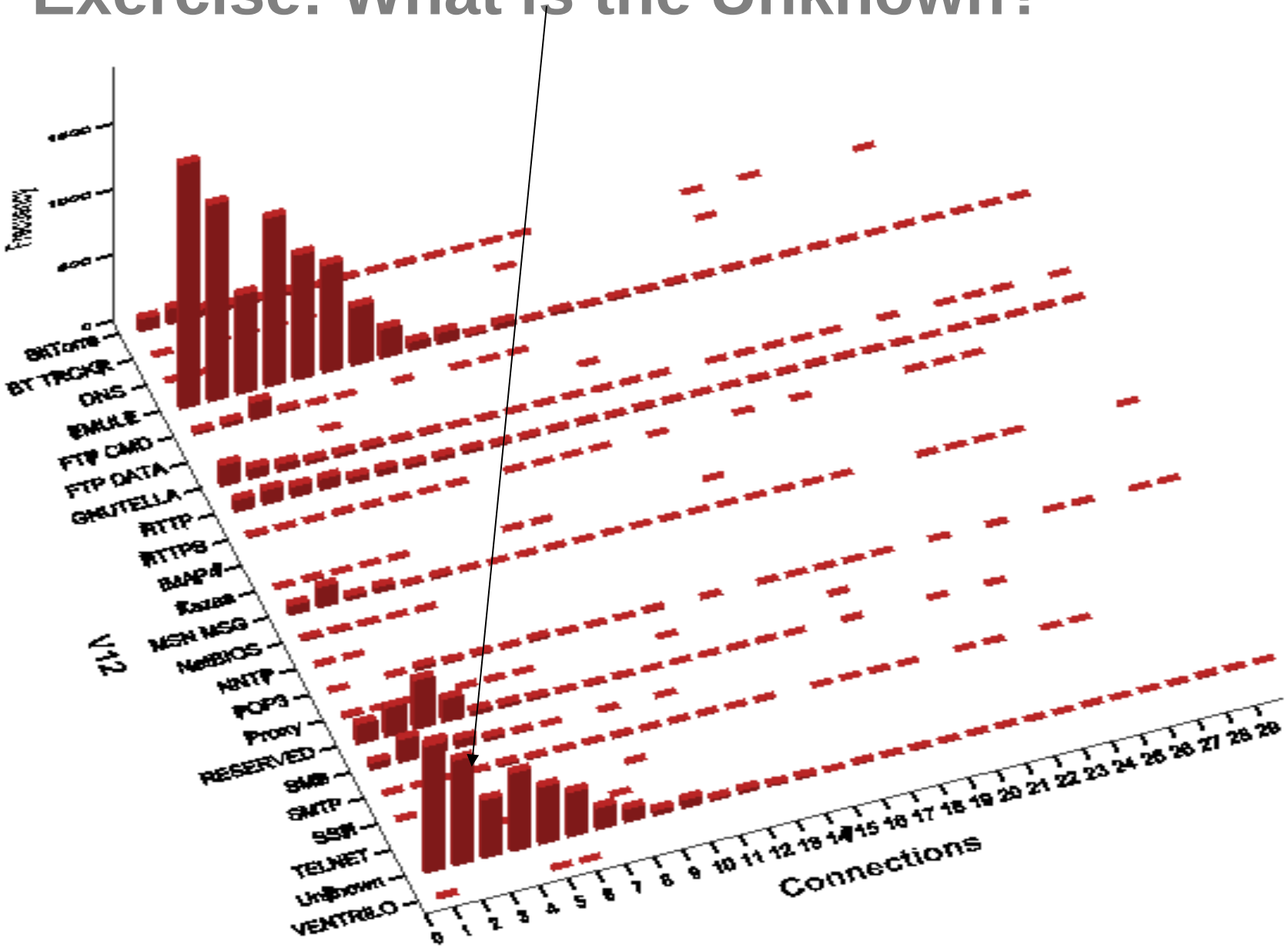


Together ahead. RUAG



Together ahead. RUAG

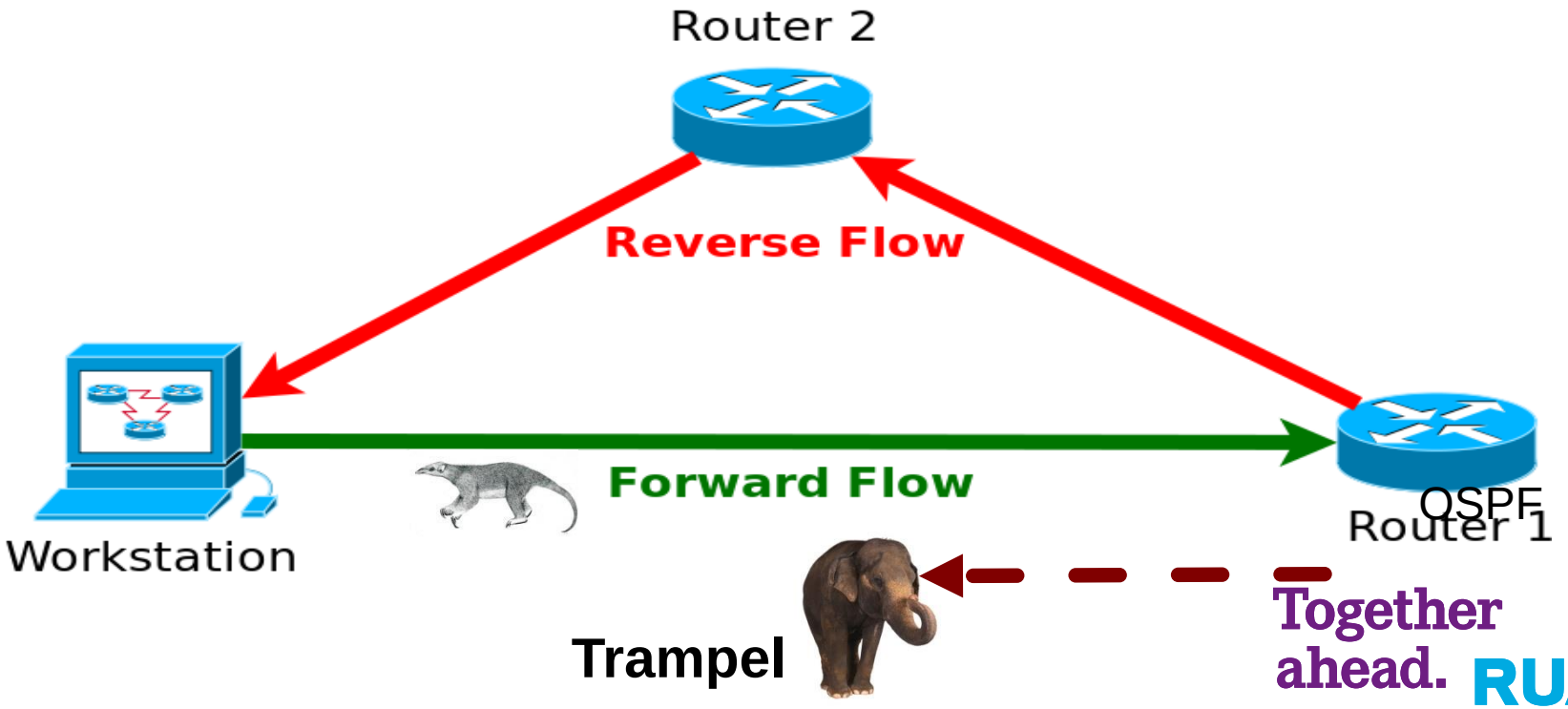
Exercise: What is the Unknown?



Together ahead. RUAG

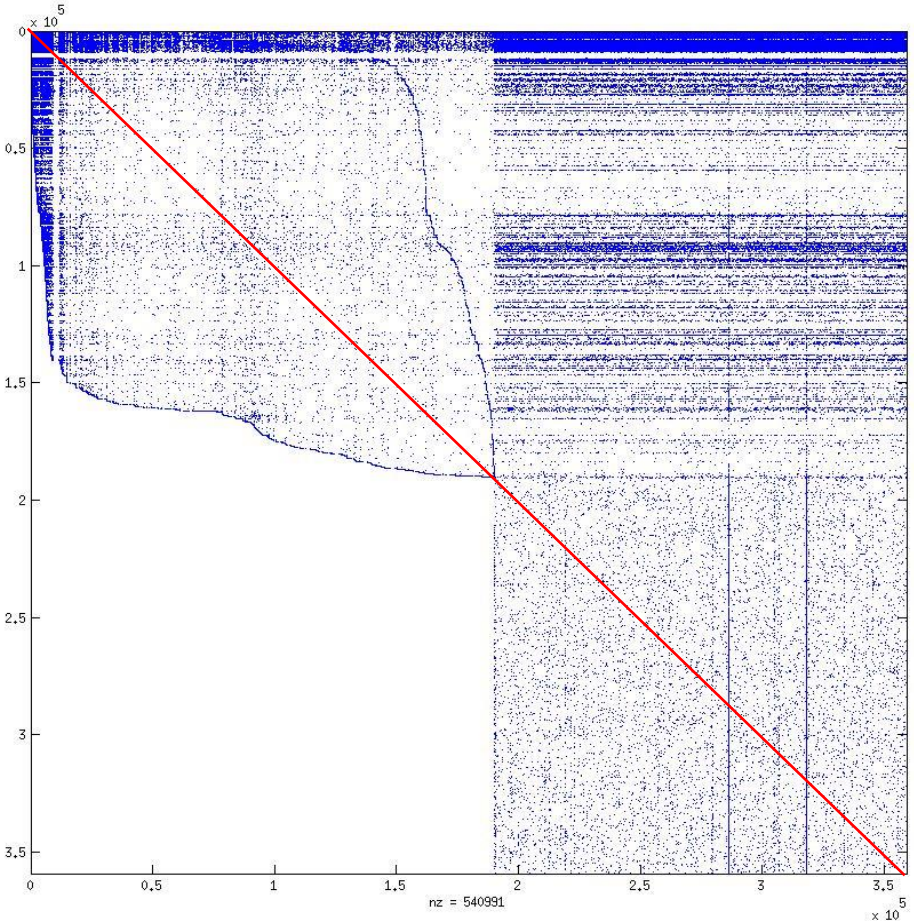
The one way TCP flow problem

Symptom: on and off access problems
TCP flows established, unidirectional
T2 proofed: Reverse connection exists, not through firewall



Together ahead. RUAG

Connection Matrix

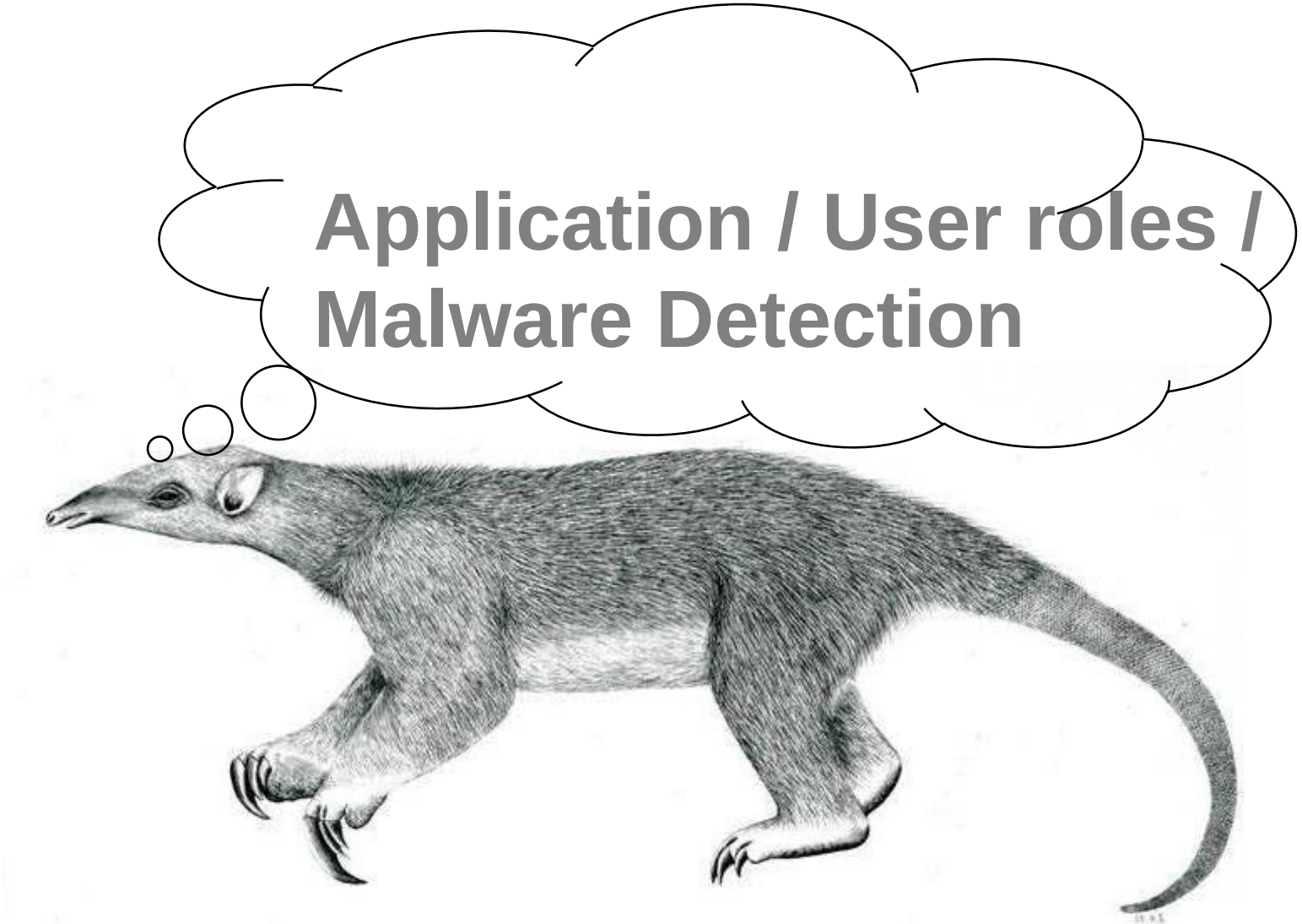


Destination IP

Source IP

Awk '{if($16) ….}' ….
Sort | uniq ….

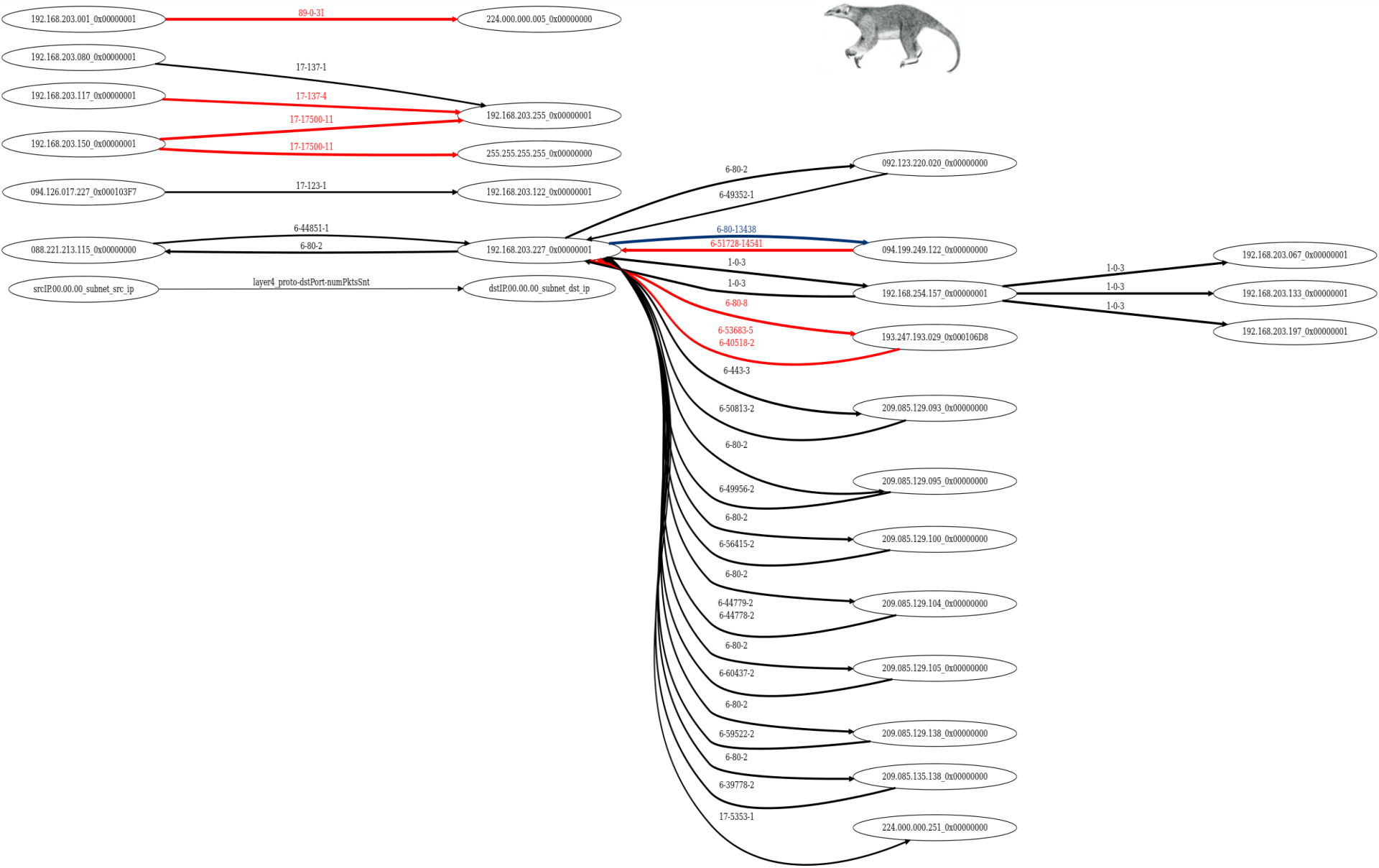
Together ahead. RUAG



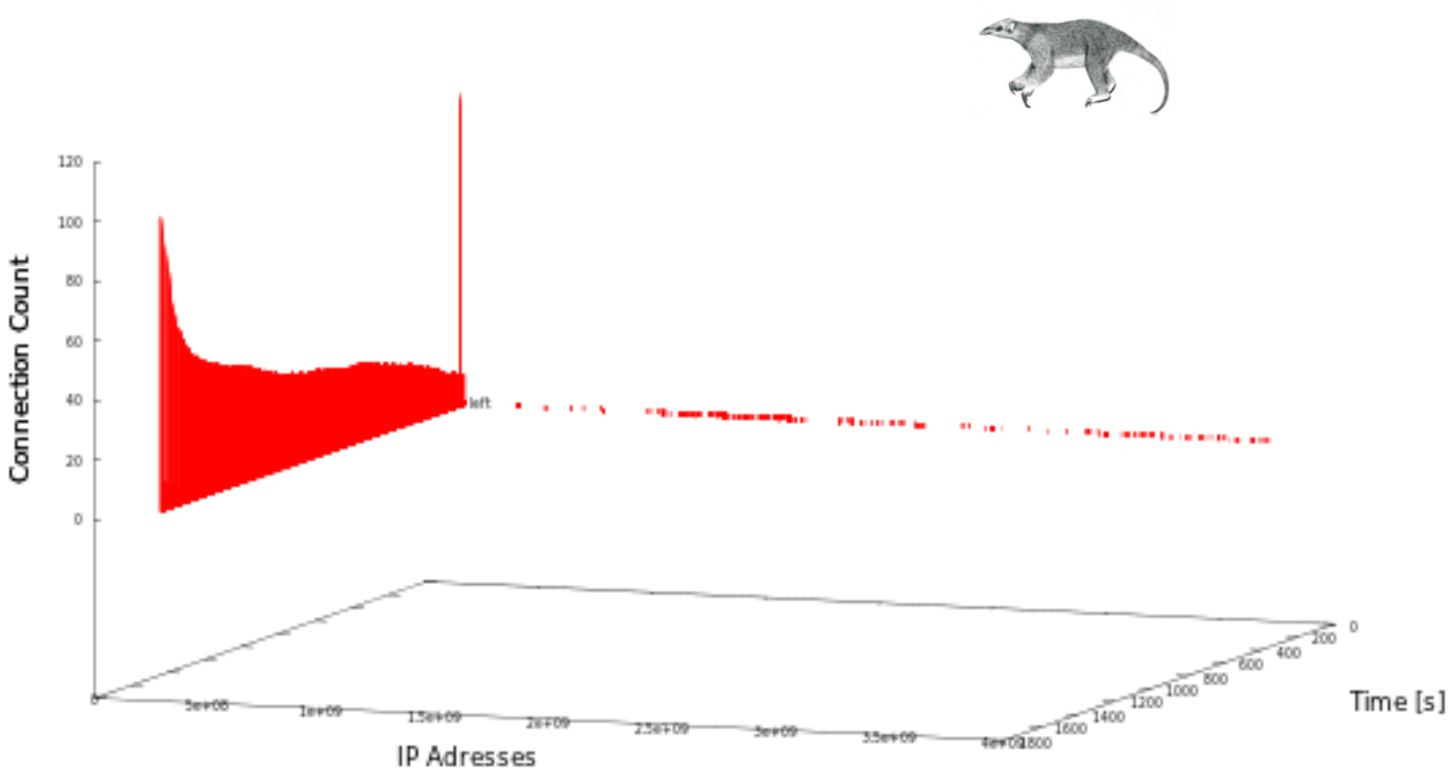
Together
ahead. RUAG

Layer3/4 Visualization

Graphviz --> simple forensic picture for small datasets



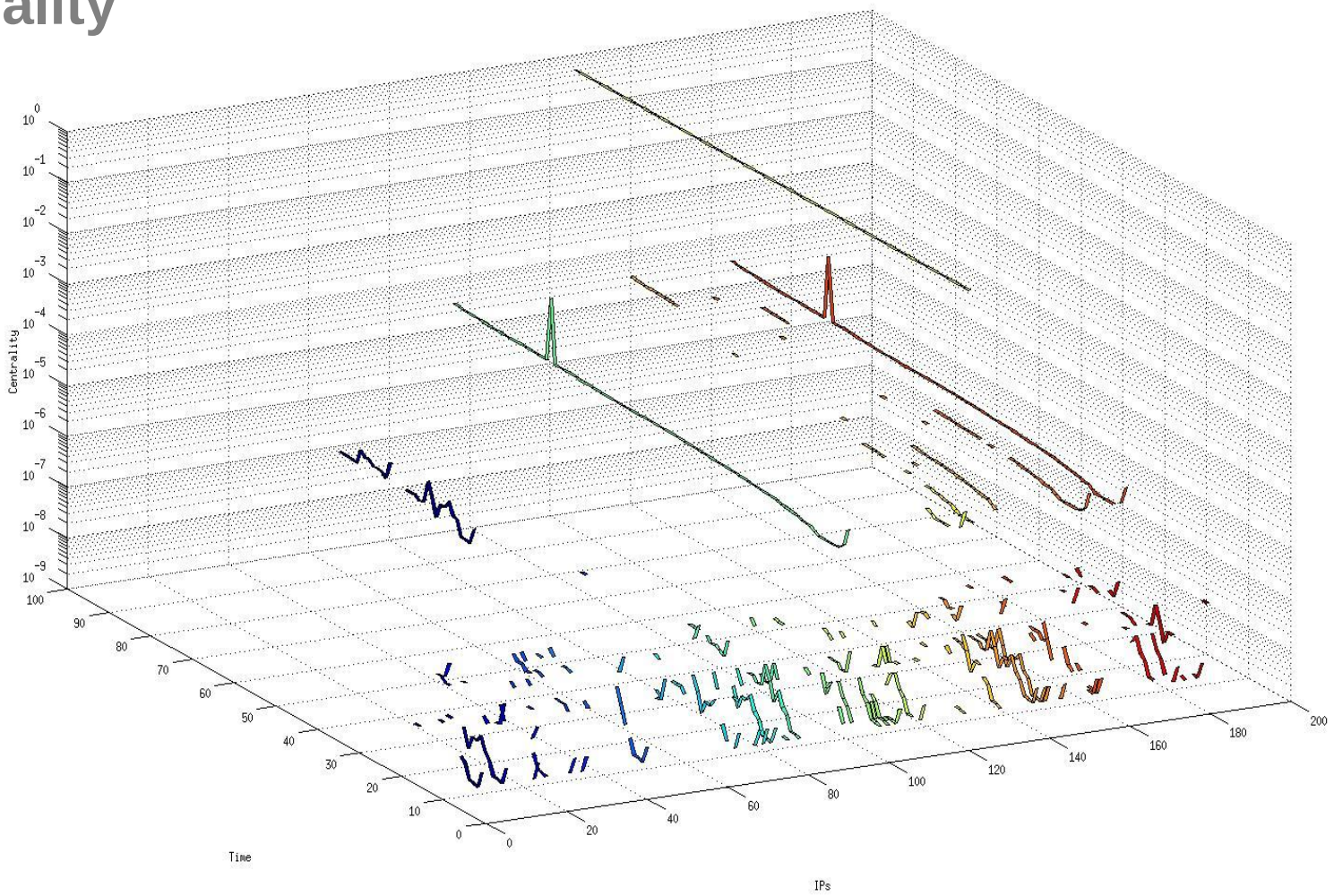
Connection Flow Matrix / t



Together ahead. RUAG

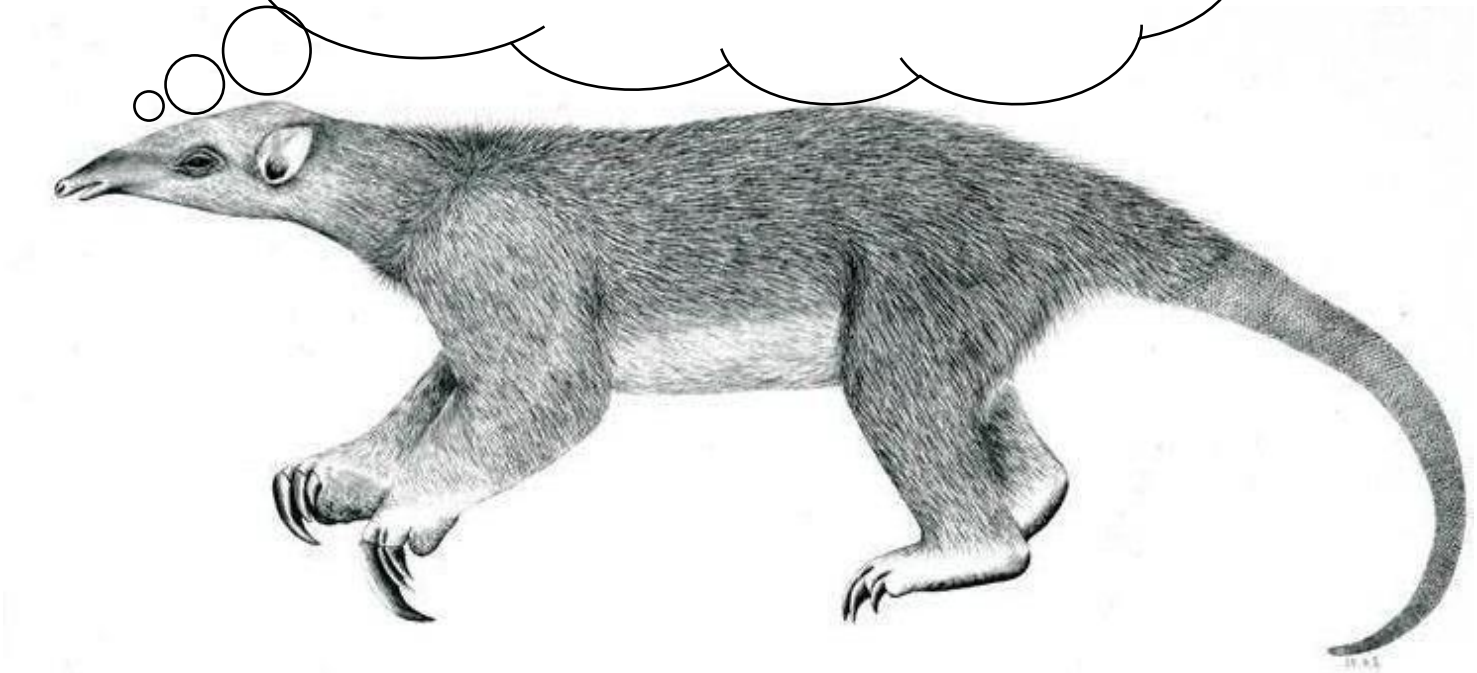
Anomaly Flow Graph: Network / Host Classification

Centrality



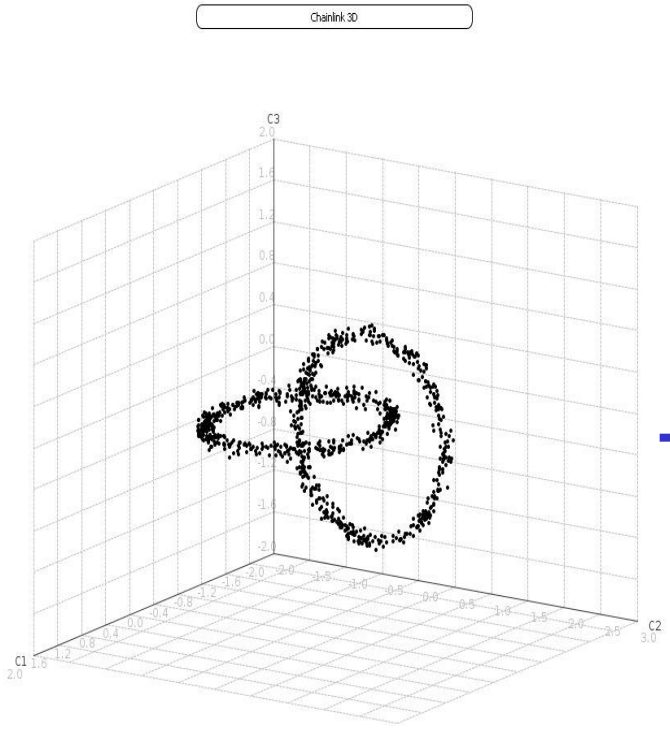
Together ahead. RUAG

Sledgehammer Dimension Reduction: AI

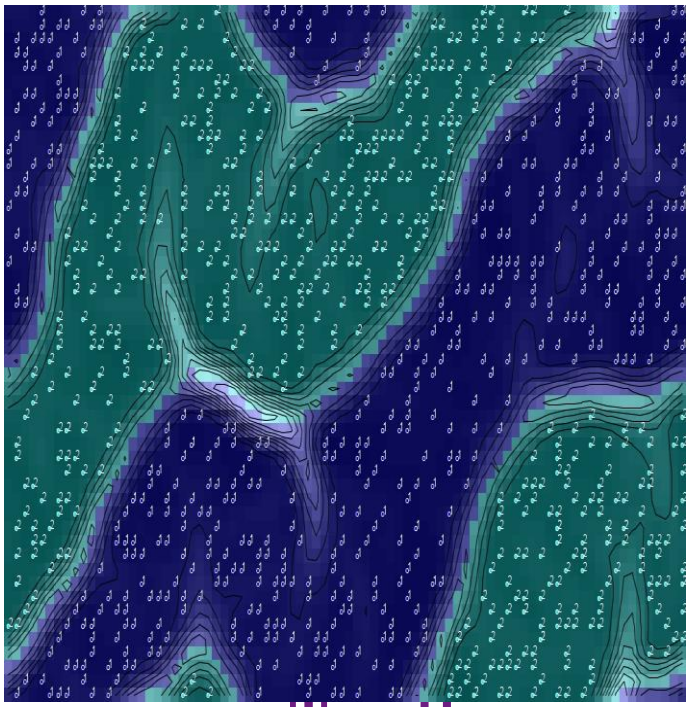


Together ahead. RUAG

Clustering of Multidimensional Data
ESOM: Nonlinear Mapping



Retina

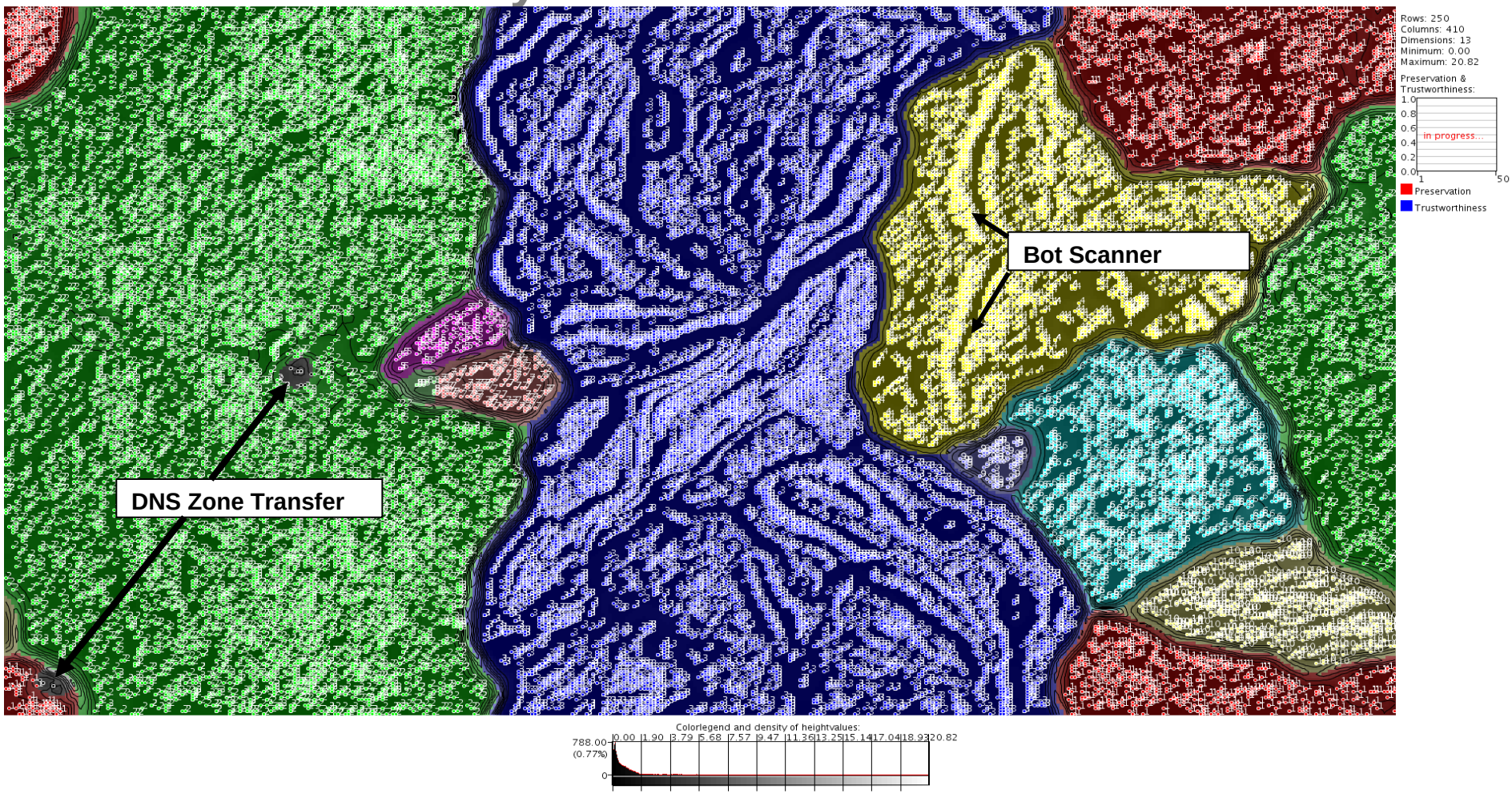


Together ahead.

ESOM Anomaly Picture:

13 Dim statistical T2 Flow parameters

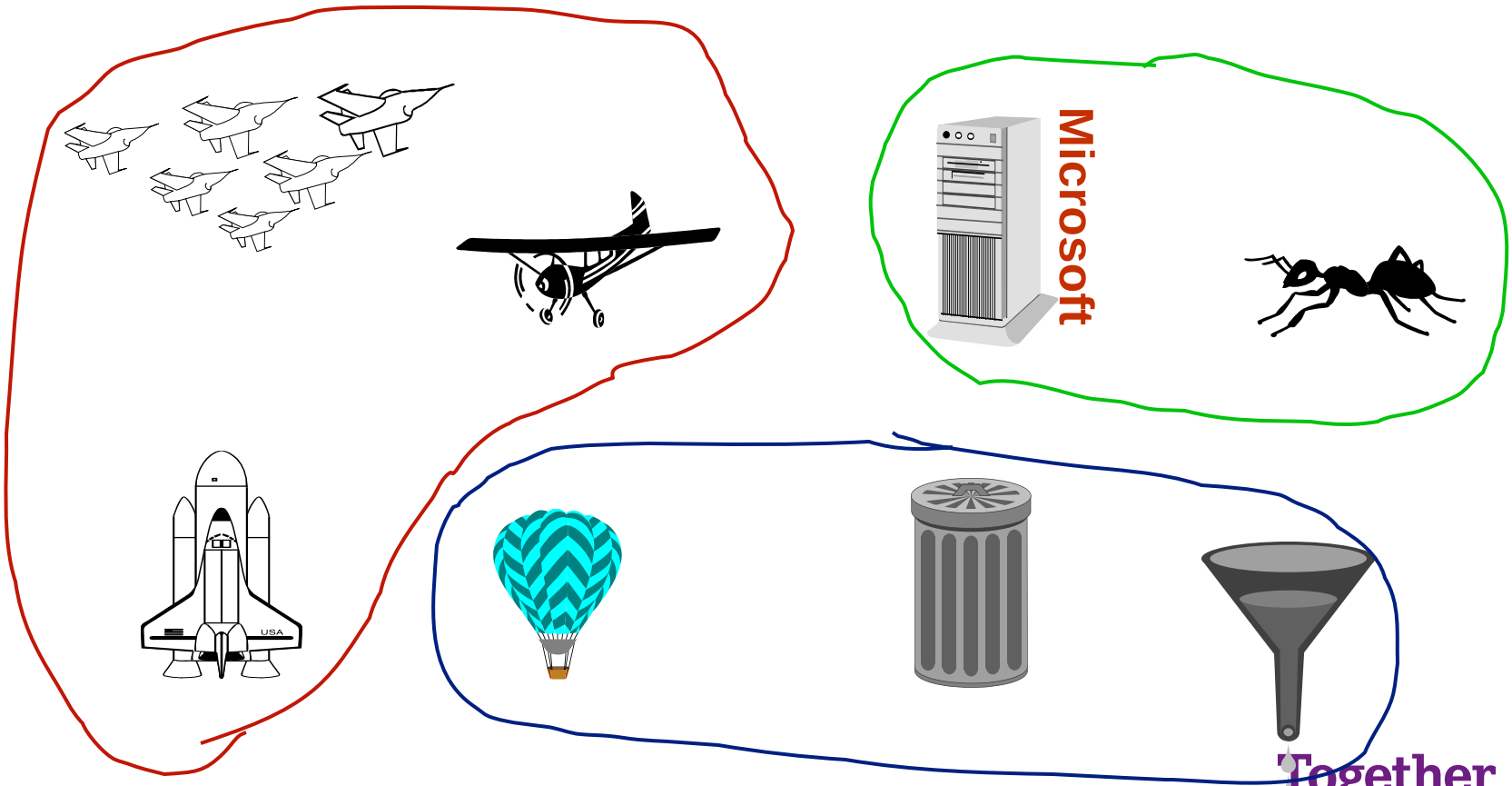
Now conceivable by human brain



Together ahead. RUAG

Together ahead. RUAG

Exercise: AI yourself



Exercise: Knowledge Extraction

Data: Bad Weather

- Low Pressure Rain Coat
- Clouds Storm
- Clouds Rain Noodle
- Sun Rain Clouds
- Spring Autumn
- Pressure sinking Heating

Data: Good Weather

- Sun Sea Bomb
- High Pressure Isobar Sun
- Pressure climb Grill Steak
- Sun Friends Beer
- Sun Beach Clouds
- Summer Sun

Good Weather = Sun & / Rain

Words or Word Chains, which separate most of data sets correctly, have highest Information gain.

Questions / Comments

Try me !
Who wants Bootcamp?

http://sourceforge.net/projects/tranalyzer/
http://tranalyzer.com
http://sourceforge.net/projects/traviz
Google: Dataming for Hackers

stefan.burschka@ruag.com

Together ahead. RUAG